þÿ T r o p i c a l s t r e a m d i a t o m c o m m u n i t i e s t h e i m p o r t
headwater streams for regional diversity

Jyrkänkallio-Mikkola, Jenny

2018

Jyrkänkallio-Mikkola , J , Siljander , M , Heikinheimo , V , Pellikka , P & Soininen , J 2018 , '
þÿ T r o p i c a l s t r e a m d i a t o m c o m m u n i t i e s t h e i m p o r t a n c e o f h e a d w a t e r s t r e
diversity ' , Ecological Indicators , vol. 95 , no. 1 , pp. 183-193 . https://doi.org/10.1016/j.ecolind.2018.07.030

http://hdl.handle.net/10138/317924
https://doi.org/10.1016/j.ecolind.2018.07.030

TROPICAL STREAM DIATOM COMMUNITIES — THE IMPORTANCE OF HEADWATER STREAMS FOR REGIONAL DIVERSITY

Jenny Jyrkänkallio-Mikkola[a], Mika Siljander[a], Vuokko Heikinheimo[a], Petri Pellikka[a] & Janne Soininen[a]

[a] Department of Geosciences and Geography, University of Helsinki, P.O. Box 64 FI-00014 Finland

*For correspondence.

E-mail jenny.jyrkankallio-mikkola@helsinki.fi; Tel. (+358) 504484125

Abstract

Understanding how species are distributed in space and time is a focal element guiding conservational efforts under the ongoing climate change and Holocene extinction. Freshwater habitats are currently one of the most threatened ecosystem types and studies aiming to unravel factors that govern biodiversity of tropical stream micro-organisms are especially scarce. Diatoms play an important role as primary producers in streams and are widely used as ecological indicators. Yet, relatively little is known about which factors affect diatom communities in the tropics.

Here, we studied benthic diatom diversity across 67 tropical streams spanning stream orders 1 – 5 in Kenya. We examined whether the hypothesis of latitudinal diversity gradient applies for benthic diatoms, i.e. do tropical streams encompass more species than boreal streams using comparable boreal dataset. In addition, we studied which environmental, land use and spatial factors control benthic diatom communities using redundancy analysis. We also examined the nestedness and turnover components of beta diversity, factors contributing to diatom species richness and the uniqueness of the communities across stream orders by using boosted regression trees and local contribution to beta diversity. Finally, we studied whether environmental heterogeneity and beta diversity are related across stream orders and tested their relationship using tests of homogeneity of dispersion and regression analysis.

Species richness was not higher in tropical streams than in boreal ones. Tropical diatom communities were controlled jointly by local environmental and spatial factors. Although water chemistry was the most important controlling factor, also physical variables contributed significantly to community variation. Land use had also a significant effect on diatom communities as broad leaved forest streams harboured different diatom communities compared to streams with higher human impact and conductivity, stressing the importance of forests to water quality and diatom biodiversity. Headwater streams encompassed highest species turnover, whereas nestedness was higher in higher order

streams. Species richness was significantly higher in higher order streams than in headwaters, whereas the uniqueness of the communities peaked in headwaters. Environmental heterogeneity was the highest in headwater streams and was related with high beta diversity, which highlights the importance of habitat heterogeneity to biodiversity. Our results stress the management and conservational importance of headwater streams and tropical montane forests as these environments harbour unique diatom communities important for regional diversity.

Keywords: beta diversity, community structure, diatoms, environmental heterogeneity, headwaters, tropics

1. Introduction

How species are distributed across Earth in space and time remains a fundamental topic guiding conservation efforts under the ongoing climate change and Holocene extinction (Pimm *et al*., 1995; Thomas *et al*., 2004). Extinction threat concerns not only terrestrial ecosystems but also aquatic communities. Freshwater habitats harbour disproportionally large species richness in relation to habitat area in many taxonomic groups and freshwater biodiversity is more threatened than biodiversity in other ecosystems (IUCN, 2009; Strayer and Dudgeon, 2010; Wiens, 2016). Thus, understanding of the drivers that maintain species diversity is urgently needed in order to combat severe biodiversity losses taking place at present.

While studies aiming to unravel the determinants of macro-organismal biodiversity have been extensive (Rosenzweig, 1995), microbial communities (e.g. bacteria, unicellular algae and fungi) have received less attention in this regard (Green *et al*., 2004; Zeglin, 2015). For example, whereas

latitudinal diversity gradients are well documented for macro-organisms (Gaston, 2000; Hillebrand, 2004; Lawton, 1999), the existence of such pattern for micro-organisms remains a controversial issue (Hillebrand and Azovsky, 2001; Passy, 2010; Salinas *et al*., 2015). Furthermore, microbial biodiversity studies in boreal and temperate regions have been far more numerous than in the tropics but see for example (Bellinger *et al*., 2006; Bere, 2014; Bojorge-Garcia *et al*., 2014; Mangadze *et al*., 2015).

The diversity and composition of microbial communities are affected by a wide range of determinants and there is still no consensus whether the key drivers are more related to the local environment (Gothe *et al*., 2013; Maloufi *et al*., 2016) or spatial factors (Crump *et al*., 2007; Soininen *et al*., 2004). Also land use (Bere and Tundisi, 2011; Carpenter and Waite, 2000) or climatic factors (Pajunen *et al*., 2016) may affect aquatic microbial communities. Finally, spatial scale of the study influences the factors which govern microbial communities (Lindstrom and Langenheder, 2012) and communities may thus often be described by the shared influence of spatial and environmental factors with increasing spatial influence in larger studies (Heino *et al*., 2014; Soininen *et al*., 2016; Verleyen *et al*., 2009).

Freshwater organisms may be especially vulnerable to changes in precipitation and possible droughts can quickly alter the freshwater habitat and thus their biodiversity (Wiens, 2016). Streams encompass disproportionately high biodiversity (Vinson and Hawkins, 1998; Vorosmarty *et al*., 2010) and stream network provides an excellent environment to study diversity patterns via its hierarchic dendritic organization (Grant *et al*., 2007). River Continuum Concept (RCC) predicts biological diversity to peak at mid-order streams, whereas headwaters and large rivers should be less diverse (Vannote, 1980). However, while this prediction might hold true for alpha (local) diversity, beta (between-site) and gamma (regional) diversity may show different patterns as especially headwater streams (orders 1–2) may harbour a great proportion of biodiversity compared to higher order stream branches (Meyer *et al*., 2007). Moreover, human induced disturbances usually increase with increasing stream order

(Januchowski-Hartley *et al*., 2011), which may affect biodiversity. Headwater streams may also be more physically variable promoting environmental heterogeneity and subsequently also beta diversity (Finn *et al*., 2011). Consequently, possible environmental homogenisation in higher order streams may decrease beta diversity compared to headwaters.

High beta diversity in headwaters may also be caused partly by the more isolated position of headwater streams compared to higher order branches (Finn *et al*., 2011). Thus, highly connected sites at lower reaches harbour lower beta diversity due to higher exchange of individuals between sites leading to homogenisation of communities (Lopes *et al*., 2014). Furthermore, anthropogenic and natural stressors may cause different patterns to beta diversity given that beta diversity under natural disturbances may be caused more by species turnover (i.e. the shared number of species between sites is small), whereas community dissimilarities under anthropogenic stressors may be more generated by nestedness (i.e. poorer assemblages are subsets of those of richer sites) (Gutierrez-Canovas *et al*., 2013). These patterns are largely unresolved in tropics, however, as stream beta diversity has not been studied extensively in tropical regions (but see (Al-Shami *et al*., 2013; Tonkin *et al*., 2016) and microbial communities have received even less attention.

Sub-Saharan Africa is going through an intense land cover change due to human population growth and subsequent cuts of forests into arable land (Brink *et al*., 2014; Pellikka *et al*. 2017a). Land cover changes can have severe effects to the environment and natural resources (Alcantara-Ayala *et al*., 2006; Hohenthal *et al*., 2015). Furthermore, land use changes in the catchments may have profound effects also on aquatic ecosystems via water scarcity, increased turbidity and nutrient fluxes (Allan and Castillo, 2007; Hohenthal *et al*., 2015; Soininen *et al*., 2015). Human impact on stream communities usually grow downstream, agricultural effects causing the most severe threat to biodiversity (US - Environmental Protection Agency, 2000; Vorosmarty *et al*., 2010). The detrimental anthropogenic effects are increased by the poor waste water treatment in many rural areas (Concoran

et al., 2010). Indeed, freshwater biodiversity conservation under the increasing human demands of water poses a true challenge (UNESCO, 2009; Wiens, 2016; Vorosmarty *et al*., 2010).

Benthic diatoms are a pivotal component of stream biodiversity and are widely used as ecological indicators (Hill *et al*., 2000; Wang *et al*., 2005). We studied the diversity and composition of stream benthic diatom communities in Taita Hills, Kenya. The area is considered as a biodiversity hotspot with a high rate of endemism of plants (Beentje and Ndiang'ui, 1988; Gereau *et al*., 2006; Lovett and Wasser, 1993) and animals (Gereau *et al*., 2006). At the same time, however, the area is affected by intensive land use change along with other anthropogenic pressures threatening ecosystems and their functioning (Pellikka *et al*., 2013; Pellikka *et al*., 2009). First (Q_1), we were interested if there are major richness differences between tropical and boreal stream diatoms. We thus compared diatom species richness between tropical and boreal streams using two data sets with identical sampling methods. We expected that species richness is higher in tropics as found for other taxa (Hillebrand, 2004). Second (Q_2), we studied the effects of local environmental, land use and spatial factors on tropical stream diatom community composition. We expected water chemistry and stream physical variables to have a significant effect on diatom composition (Soininen, 2007) but also that communities are spatially structured (Piano *et al*., 2017). In addition, we expected land use to affect diatom communities indirectly via affecting water quality (Bere and Tundisi, 2011; Carpenter and Waite, 2000; Potapova and Charles, 2002). We considered the use of the diatoms and their diversity as ecological indicators of land cover change by sampling diatoms in four massifs of the Taita Hills of Kenya, which differ in their land use history, but are similar in their bedrock, soil types, forest species composition, agricultural crop composition and climate. Thirdly (Q_3), we studied the distribution of diatom biodiversity across stream orders. We expected headwater sites (orders 1–2) to harbour more unique diatom communities, and simultaneously, to have lower species richness than higher stream order sites (Meyer *et al*., 2007; Sherwood *et al*., 2000). We also examined beta diversity components, i.e., species turnover and nestedness across stream orders and expected species turnover

to be largest in headwaters while nestedness to be largest in downstream sites. Fourthly (Q_4), we examined factors contributing to diatom species richness and the uniqueness of the communities. We expected that species richness is mostly related to conductivity (Heino *et al*., 2010) and pH (Jyrkänkallio-Mikkola *et al*., 2017) whereas uniqueness would be most related to conductivity and nutrient levels (Pajunen *et al*., 2017). Finally (Q_5), we studied whether environmental heterogeneity and diatom beta diversity are related across stream orders and tested their relationship. We expected headwater streams to harbour higher beta diversity and environmental heterogeneity (Finn *et al*., 2011; Meyer *et al*., 2007) than sites at higher order streams.

In this study, we characterized diatom communities using species richness, community composition and the uniqueness of species composition at sites, measured as local contribution to beta diversity (LCBD; (Legendre and De Caceres, 2013). Since beta diversity provides the most useful measure to understand the processes that generate and maintain biodiversity (Legendre *et al*., 2005; Qiao *et al*., 2015; Socolar *et al*., 2016) we placed most emphasis on beta (between site) diversity in terms of dissimilarities and local contribution to beta diversity.

2. Material and methods

2.1. Study area

Taita Hills is the northernmost part of the biodiversity hotspot Eastern Arc located in southeastern Kenya rising in the middle of plains to hilltops over 2000 m of an altitude (Platts *et al*., 2011). Taita Hills can be separated to Dabida massif, Mbololo massif, Sagala Hill and Mt. Kasigau, which are within 40 km from each other. The mean elevation of Dabida is between 1300 to 1600 m a.s.l., the highest peaks reaching over 2000 m a.s.l. Mbololo massif reaching 1800 m is located east from Dabida separated by Paranga valley at 900 m a.s.l. Sagala Hill reaching 1500 m a.s.l. is located 15 km southeast from Dabida and Mbololo, and Mt. Kasigau reaching 1600 m a.s.l. at 40 km southeast

156 from Dabida (Fig. 1). Located in the intertropical convergence zone Taita Hills face bimodal rain
157 pattern having rainy season in November-December and in March-May. Annual rainfall in the hills
158 is normally 1200 mm, while the lowlands receive less than 600 mm per year resulting to rule of the
159 thumb that higher elevations have higher rainfall and are covered more by forests than lower
160 elevations (Erdogan *et al*., 2011). However, as the rains are orographic in nature, the eastern slopes
161 and hills of Mbololo and Dabida receive more precipitation than the leeward parts and slopes located
162 in the west. A significant addition to the precipitation is occult precipitation, which is estimated to
163 result in 20% addition to precipitation, but only in the forests as forest volume is required to capture
164 the atmospheric moisture, the mist (Pellikka *et al*, 2017b).

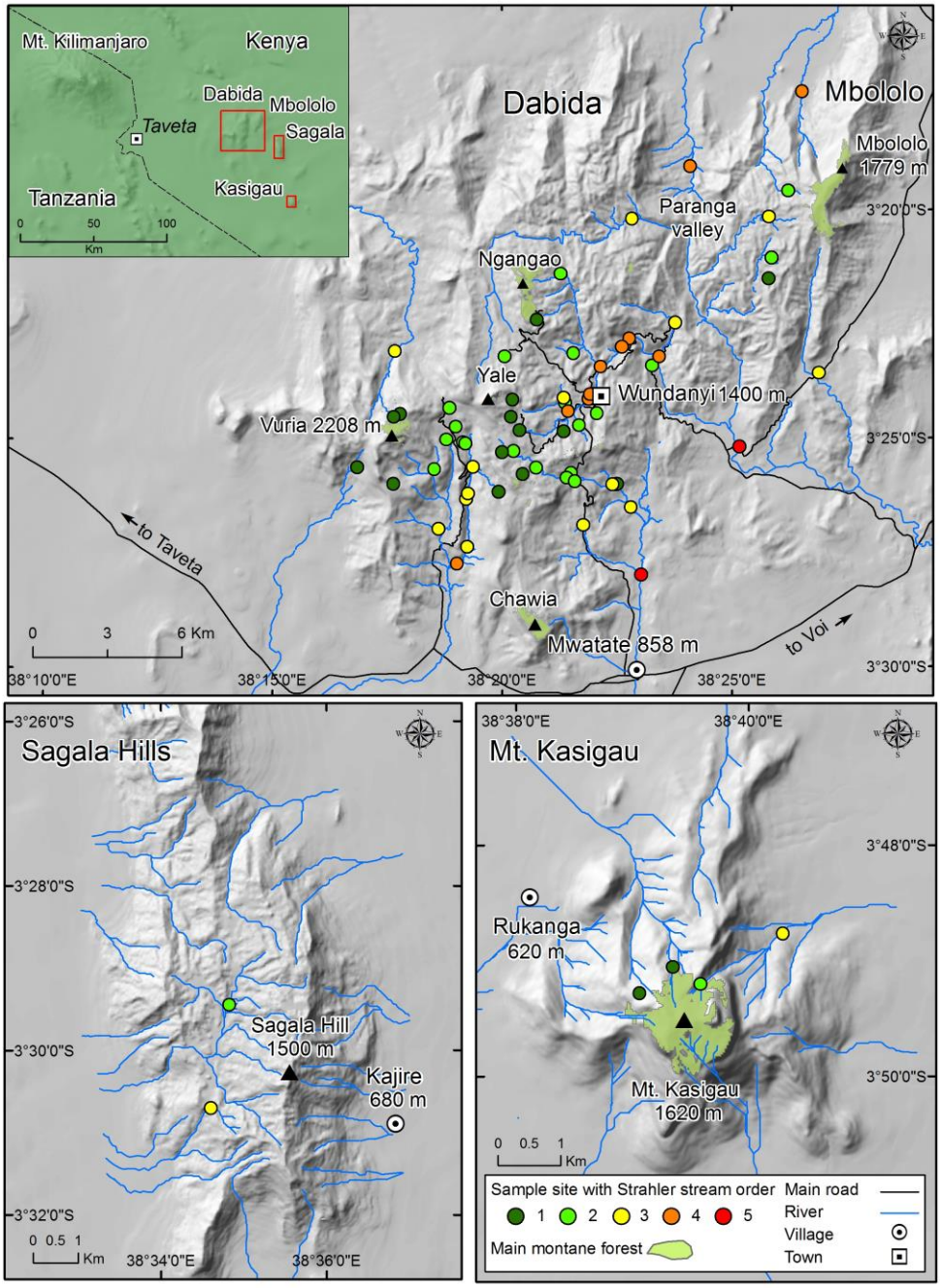


Figure 1 Map of the sampling area

It is estimated that the hills over approximately 1220 m a.s.l. prior human disturbance were covered by forests characterized as evergreen moist lower or upper montane forests. Now only a few percent of the montane forests are left mainly as protected forests managed by Kenya Forest Service. Below 1220 m a.s.l. the forests can be characterized as Acacia-Commiphora woodland, which is partly deciduous depending on the species. The hills are cleared for intensive small-scale agriculture the

main crop being maize producing to harvest per year (Jaetzold and Schmidt, 1987). The lowlands are used for dryland farming, grazing by goats and camels, sisal plantations and conservation areas.

The four massifs studied differ in their land use. Based on Pellikka *et al*. 2017a, only the hilltops of Dabida and Mbololo massifs are covered by montane forests. Due to cloudy conditions, the peak of Vuria (2208 m a.s.l.) is characterized as upper montane forest, elfin forest, with trees covered heavily by epiphytes (Stam *et al*., 2017), while the other native forests are lower montane forests with less epiphytes, but trees reaching up to 40 m of a height. Over 1220 m a.s.l. in 241 km^2 of an area, croplands cover nowadays 40.4 %, montane forests 3.2 % and plantation forests of mainly pine and eucalyptus 10.1%, and woodland, shrubland and thickets all together 44.8% (Pellikka *et al*. 2017a). The remaining fragments of montane forests in Dabida and Mbololo massifs are 1–2 km^2 in size: Mbololo (185 ha), Ngangao (120 ha), Vuria (ha) and Chawia (86 ha) being the largest ones. These forests are all degraded due to selective timber harvesting and grazing inside of the forest, especially in Vuria and Chawia. The least degraded is Mbololo forest and then Ngangao (Omoro *et al*., 2013).

Sagala hill has no native montane forests left, but has some plantation forests and degraded woodlands. The montane forest of Mt. Kasigau can be characterized as upper montane forest over 1500 m a.s.l. due to misty conditions producing fog deposit, and lower montane forest down to 1000 m a.s.l. (Medley and Maingi, 2014) below which dryland forest like Acacia-Commiphora with Euphorbia ssp. takes place. The montane forest of Kasigau is least disturbed of all the forests of the Taita Hills being 2.03 km^2 in size (Medley and Maingi, 2014). Croplands did exist in Mt. Kasigau few decades ago, but were abandoned in order to safeguard water resources (Rikkinen *et al*., 2015). Evidently due to very steep slopes they were also difficult to manage (Adhikari *et al*. submitted).

2.2. Biological data collection

Diatom samples were collected at 67 sites from all four mountain massifs (Fig.1). Sampling was completed in three weeks in January 2016. Diatoms were sampled by randomly selecting and brushing 10 cobble-sized stones or bedrock. The samples were collected from different part of the stream site in order to control small-scale variability in the community composition. To standardize the sampling area we used a 5 × 5 cm rubber template. The samples were preserved with 99% ethanol and stored at +4 °C. Wet combustion with hydrogen peroxide (20%, H_2O_2) was used to clean diatom frustules from organic material. Subsequently, samples were mounted on microscope slides using high-refractive mounting resin Naphrax. We counted and identified at least 500 frustules from each sample using phase contrast light microscope with 1000 × magnification. Species identification was based on Krammer and Lange-Bertalot (1986-1991) and Taylor *et al*. (2007).

2.3. Chemical and physical measurements

Water samples for total nitrogen, total phosphorus and water colour were collected a few meters upstream from each sampling site. Total nitrogen was determined by using alkaline persulfate digestion where alkaline oxidation is carried out in a thermo digester and the converted total nitrogen quantified colorimetrically. Total phosphorus was determined by inductively coupled plasma-optical emission spectrometry (ICP-OES). However, as the method lacked the sensitivity to determine phosphorus concentrations below 10 µg L^{-1} and the phosphorus content of 51 samples were ≤ 10 µg L^{-1} we were unable to use total phosphorus in the following analyses. Water colour was analysed using standardized method (EN ISO7997).

To measure conductivity, pH and temperature in the field we used YSI professional plus water quality meter (YSI Incorporated, Yellow Springs, USA). Current velocity was measured with Schiltknecht MiniAir 2 flow meter (Schiltknecht, Gossau, Switzerland) from 30 locations covering the sampling

site extensively. Water depth was similarly measured from 30 locations and stream width from 10 locations covering the length of the whole sampling site.

Substratum particle size was measured from 10 locations using a 50 × 50 cm quadrat placed in random sites on the streambed. We used a modified Wentworth scale to visually classify the substratum particle size: sand (diameter 0.25 mm – 2 mm), gravel (2 mm – 16 mm), pebble (16 mm – 64 mm), cobble (64 mm – 256 mm) and boulder (>256 mm) (Wentworth, 1922). Canopy shading was individually estimated from 20 locations covering the whole sampling site. Range and average values of measured environmental variables are listed in Appendix A in the supplementary material.

2.4. Stream order analyses and land cover data

River network and stream ordering were calculated from Digital Elevation Model (DEM) data. DEM of 1 × 1 m resolution was derived from airborne laser scanning (ALS) data sets for two parts of the study area (Taita Hills and Kasigau). Laser scanning data was acquired in 2013 and 2014 (using Optech ALTM 3100), and data was filtered for buildings, powerlines and outliers using Terrascan software (Terrasolid Oy), LasTools (Rapidlasso GmbH) and manual editing. For Sagala area, a DEM was created from scanned Survey of Kenya 1:50 000 scale topographic map from which the contour lines were digitized and converted to 10 m resolution raster DEM.

Watershed delineation was based on the DEM layers and, hydrological modelling tools in ArcGIS 10.3.1. Firstly, elevation grids were filled to create a surface that has no sinks. Then, flow direction grid was calculated to determine the direction of flow for each cell in the landscape. Flow accumulation function was used to generate a drainage network for the selected pour points. The Snap Pour Point tool was used to ensure selection of points of high accumulated flow and finally, watersheds were automatically delineated. ArcGIS 10.3.1 hydrological tools and Stream Order function with Strahler method (Strahler, 1957) were used to determine stream ordering from the river

networks. Elevation models were harmonized by resampling the DEM layers into 10 m pixel size. River networks were then determined again and stream ordering calculated.

Land cover was assessed from classifies satellite imagery separately for Taita Hills, Kasigau and Sagala. For Taita Hills, the land cover classification was based on a 20 × 20 m resolution SPOT 4 satellite image from 23 October 2011 (CNES, 2013). The satellite image was classified into 9 land cover classes (Pellikka *et al*., 2017; Appendix A in the supplementary material) using an object-based approach (Benz *et al*., 2004) following the methodology presented in Clark and Pellikka (2009). Segmentation and classification was implemented in the eCognition software (Trimble). After manual corrections, the overall accuracy of the land cover map was 71.1% when comparing the map with 305 random ground-truth points.

For Sagala area, Sentinel-2A MSI Level-1C satellite image from 8 October 2016, downloaded from the Sentinel's Scientific DataHub (ESA, 2015) and RandomForest (Breiman, 2001) classifier in R software were used to create the land cover classification. Firstly, Sentinel-2A was DOS1 atmospheric corrected using Semi-Automatic Classification Plugin in QGIS. Then, classifier training areas were on-screen digitized from very high-resolution (1 m pixel size) image mosaic from 2004 and from more up to date satellite imagery (Google Earth, 2016). Then, reflectance values from the four bands blue, green, red and infra-red from the Sentinel 2A satellite image were used as the predictors for RandomForest classifier. The image was classified to the following relevant classes: bush, fields, forest and impervious.

For the Kasigau area, information from the same Sentinel 2A satellite image was combined with canopy height model (CHM) and DEM derived from the ALS data for creating the land cover classification map. First, the CHM was reclassified to 5 elevation classes in ArcGIS 10.3.1. NDVI (normalized difference vegetation index) values derived from the Sentinel-2A satellite image were used to separate non-vegetated areas from different type of green areas. Then, Kasigau study area was classified into highland and lowland zones based on the 1 m pixel size DEM to separate montane

267 vegetation from lowland vegetation after which ArcGIS Map Algebra tool was used to derive land
268 cover classes listed in Appendix A in the supplementary material. The resulting land cover map was
269 evaluated using GPS field measurements, very high resolution airborne imagery (1 m pixel size) and
270 also the previous knowledge of the area. Prior to statistical analyses, the land cover classes of the 6
271 sampling sites from Sagala and Kasigau were equated with those classified from Taita Hills. Finally,
272 zonal statistics of the percentages of land use and land cover classes were calculated for different
273 watershed areas based on the land cover layers and delineated watersheds for the whole study region.

274

275 **2.5. Statistical analyses**

276 Prior to analyses, explanatory variables measured as percentages were arcsine-transformed while
277 other explanatory variables were log-transformed. In addition, diatom abundance data were Hellinger
278 transformed to better meet the requirements of linear ordination methods (Legendre and Gallagher,
279 2001). Correlations between explanatory variables were tested using Pearson correlation. As none of
280 the correlations exceeded 0.70, we kept all the explanatory variables in the subsequent models
281 (Dormann *et al*., 2013).

282 Failure to notice spatial autocorrelation can lead to false interpretations of environmental variables'
283 effects on community composition. Thus, spatial autocorrelation of the local environmental variables
284 and species richness was tested using function correlog in the R package 'pgirmess' (Giraudoux,
285 2015). The significance of the correlogram was tested at level $P \leq 0.05$ with applied Bonferroni
286 correction (P/k, where k is the number of distance classes used). Here, distance class of 7 km was
287 used as it was most suitable for the spatial distances in our data.

288 To test whether more species are found from the tropics than boreal areas (Q_1), species accumulation
289 curves (function specaccum in the R package 'vegan′; (Colwell and Coddington, 1994; Palmer,
290 1990)) were used. The used boreal stream diatom data are explained in detail in Jyrkänkallio-Mikkola

291 *et al*. (2017). To standardize the number of sampling sites, 67 samples from the boreal data originally
292 comprising 105 sites were chosen. First order Jackknife method was used to estimate the total species
293 richness for both biomes.

294 To model spatial structures among tropical study sites, distance-based Moran´s eigenvector maps
295 (dbMEM) derived from spatial coordinates (function dbmem in the R package 'adespatial'; (Dray *et*
296 *al*., 2017)) were used. Threshold value was calculated using function give.thresh. Only eigenvectors
297 with positive autocorrelation were used to model spatial effects in subsequent analyses.

298 Redundancy Analysis (RDA) was applied to study community structure (abundance data)-
299 environment (excluding land use) relationships (Q_2) for the whole tropical dataset (Legendre and
300 Legendre, 2012). Another RDA was conducted exclusively for 39 independent sites (i.e., sites that
301 had independent catchments) to examine the effects of land use and catchment size on the diatom
302 communities. Stepwise model (forward selection) with 200 permutations was used to create final
303 RDA models (function ordiR2step in the R package 'vegan'; (Oksanen *et al*., (2015)).

304 Distribution of biodiversity across stream orders (Q_3) was studied with a set of analyses. To study the
305 uniqueness of diatom community composition across sites, local contribution to beta diversity
306 (LCBD) was calculated for each site (function beta.div in the R package 'adespatial'; (Legendre and
307 De Caceres, 2013)). The analysis was based on abundance data for which Hellinger dissimilarity
308 coefficient was applied and the analysis was run with 999 permutations. Linear regression was applied
309 to study the relationship between species richness and LCBD across sites and how richness and LCBD
310 varied across stream orders. To study whether different dissimilarity components govern diatom beta
311 diversity across stream orders, species turnover, nestedness and total beta diversity were calculated
312 for two groups of stream orders, 1-2 and 3-5, respectively (function beta.multi in the R package
313 'betapart' (Baselga *et al*., 2017)).

To explain variation in species richness and LCBD using environmental variables (Q_4), boosted regression trees (BRT; function gbm in R package 'gbm', (Elith *et al*., 2008)) were used. Tree complexity was set to 3, learning rate to 0.001 and bagging fraction to 0.75. The benefits of this machine learning technique include, that it can be run with raw values, it takes into account interactions between predictors and has small prediction errors (Elith *et al*., 2008).

Test of homogeneity of dispersion (PERMDISP, function betadisper in the R package 'vegan' (Anderson, 2006; Anderson *et al*., 2006)) was applied to relate environmental heterogeneity (habitat variables) within stream orders with corresponding variation in diatom community composition (Q_5). Sørensen coefficient was used for presence-absence data and Euclidean distances were applied for standardized habitat variables. PERMDISP tests among group differences from individual observation to their group centroid using ANOVA F-statistics (Anderson *et al*., 2006). Null hypothesis that there is no difference in diatom beta diversity and environmental heterogeneity among the different stream orders was tested. All tests were run using 999 permutations. Finally, null hypothesis that there is no relationship between the degree of beta diversity and the degree of environmental heterogeneity across the different stream orders was tested using linear regression.

Explained variation in RDA and linear regressions were reported using adjusted R^2- values. All analyses were conducted in R version 3.2.2 (R Development Core Team, 2013) using packages ´adespatial' (Dray *et al*., 2017), ´betapart´ (Baselga *et al*., 2017), 'corrplot' (Wei and Simko, 2016), 'dismo' (Hijmans, 2015), 'gbm' (Ridgeway, 2013), 'pgirmess' (Giraudoux, 2015) and 'vegan' (Oksanen *et al*., 2015).

3. Results

Altogether 297 taxa were identified from 67 tropical sites, of which 104 were found from only one site. Local diatom species richness varied from 15 to 71 in the tropical data set. The random selection

of 67 sampling sites from comparative boreal dataset resulted in taxa richness of 303. Species accumulation curves revealed that there was only a minor difference in the proportion of total species richness for boreal (76.4%) and tropical (74.6%) datasets (Q_1; Appendix B in the supplementary material).

Out of local environmental variables, only water temperature and stream width showed significant spatial autocorrelation on short distances (Appendix C in the supplementary material). Species richness showed no spatial autocorrelation. dbMEM returned one spatially autocorrelated eigenvector (referred henceforth as MEM1), that was used to model spatial structures among sites in the subsequent RDA and BRT analyses.

In the first RDA (67 sites), forward selection method returned seven significant ($P < 0.05$) variables which together explained 15.6% of the total community variation (Q_2, Fig.2a). Conductivity and pH were among the most important variables for diatom composition. Also the physical variables (shading and gravel), spatial (elevation and MEM1) and stream order were among the significant variables structuring the community. The first RDA-axis divided the sites with high conductivity, pH and high stream order opposite to those with high elevation. The second RDA-axis was mostly related to shading. We further tested whether removing singletons would increase the explained variation in community composition. This resulted only to minor increase (15.9%) of the explained variation with the same explaining variables. When presence-absence data were used explanatory power decreased to 10.9%. In the second RDA (39 sites) including land use variables, selection method returned four ($P < 0.05$) significant variables which explained 14.4% of the total community variation (Q_2, Fig. 2b). The first RDA-axis divided the sites to those with high percent of broad leaved forest and high elevation opposite to those with high conductivity.

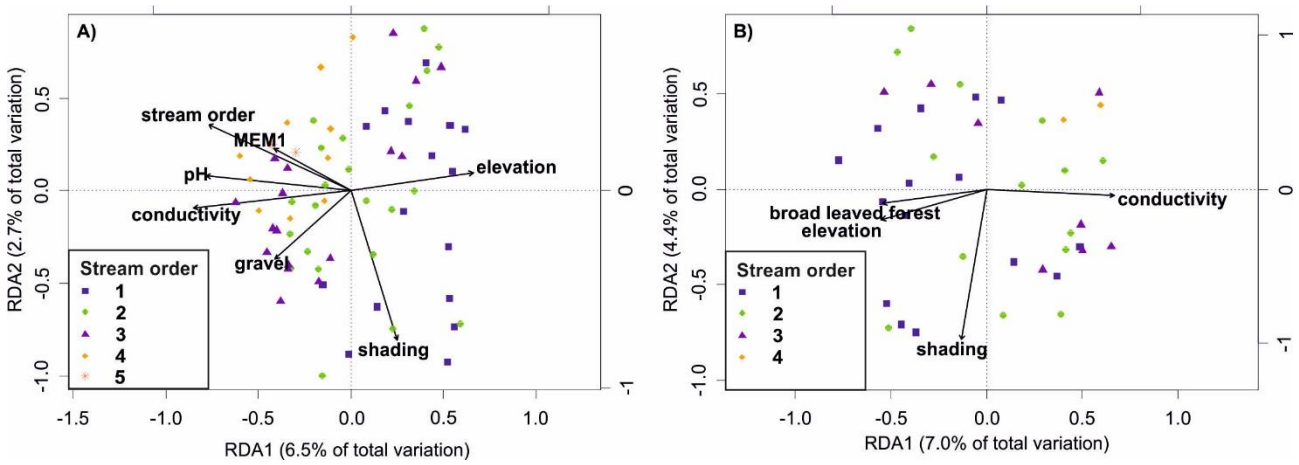


Figure 2 Ordination plots for redundancy analysis for A) all 67 sampling sites and environmental variables and B) for 39 sites with independent catchment areas and environmental and land use variables. Both ordination plots are based on Hellinger transformed diatom abundance data. Arrows represent the significant environmental variables. The explained variation is based on adjusted R^2. Abbreviations: MEM1 (spatial eigenvector derived from distance-based Moran´s eigenvector maps)

According to linear regression, species richness increased significantly with increasing stream order (Q_3, Fig.3a, $P < 0.001$, $R^2 = 0.142$), whereas LCBD values showed an opposing trend with the highest values in headwater streams (Q_3, Fig.3b, $P < 0.05$, $R^2 = 0.053$). LCBD and species richness showed significant negative relationship (Q_3, Fig. 3c; $P < 0.001$, $R^2 = 0.425$). Total beta diversity was higher in headwater streams (Q_3, orders 1-2, $\beta_{sor} = 0.933$) than in higher order streams (orders 3-5, $\beta_{sor} = 0.867$). Beta diversity components varied across stream orders. As we expected, species turnover was higher in headwater streams (orders 1-2, $\beta_{sim} = 0.905$) than in higher order streams (orders 3-5, $\beta_{sim} = 0.860$), whereas nestedness was lower in headwater streams ($\beta_{nes} = 0.027$) than in higher order streams ($\beta_{nes} = 0.036$).

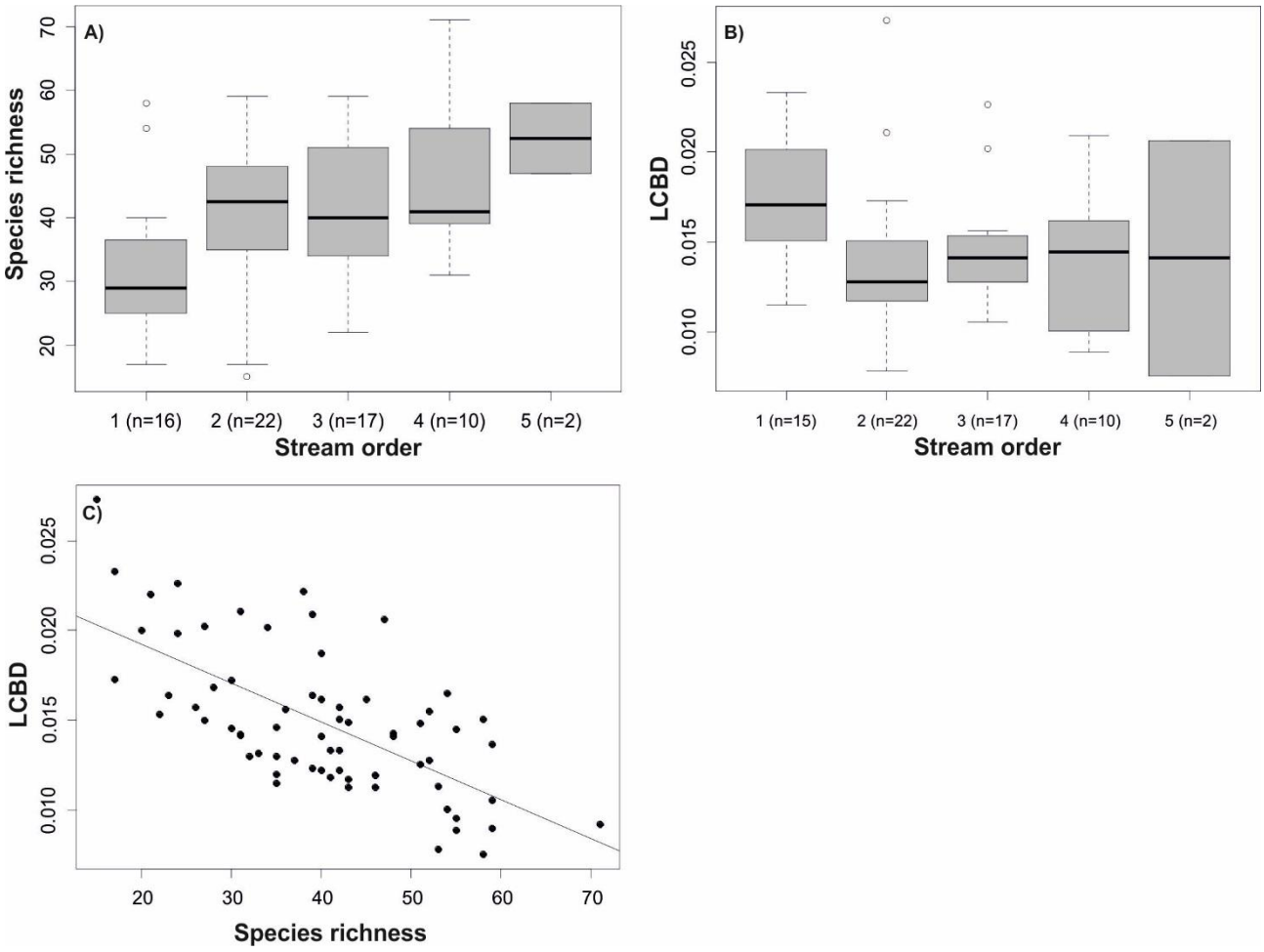


Figure 3 Boxplots representing the relationship between A) species richness and B) LCBD values and stream orders. Also shown is C) the linear relationship between LCBD and species richness. Abbreviations: LCBD (local contribution to beta diversity).

In BRT, the most important factors affecting species richness were pH, temperature, conductivity and spatial variable (MEM1) all of which showed positive relationship with species richness (Q_4, Fig.4a). Here, we also used polynomial regression and found that the relationship between species richness and temperature was unimodal ($P < 0.01$, Adj. $R^2 = 0.148$). According to BRT, the most important variables affecting LCBD were conductivity, temperature, water color and stream velocity (Q_4, Fig.4b). Conductivity, temperature and, pH were negatively related with LCBD, whereas water color scaled positively with LCBD (Fig. 4b).

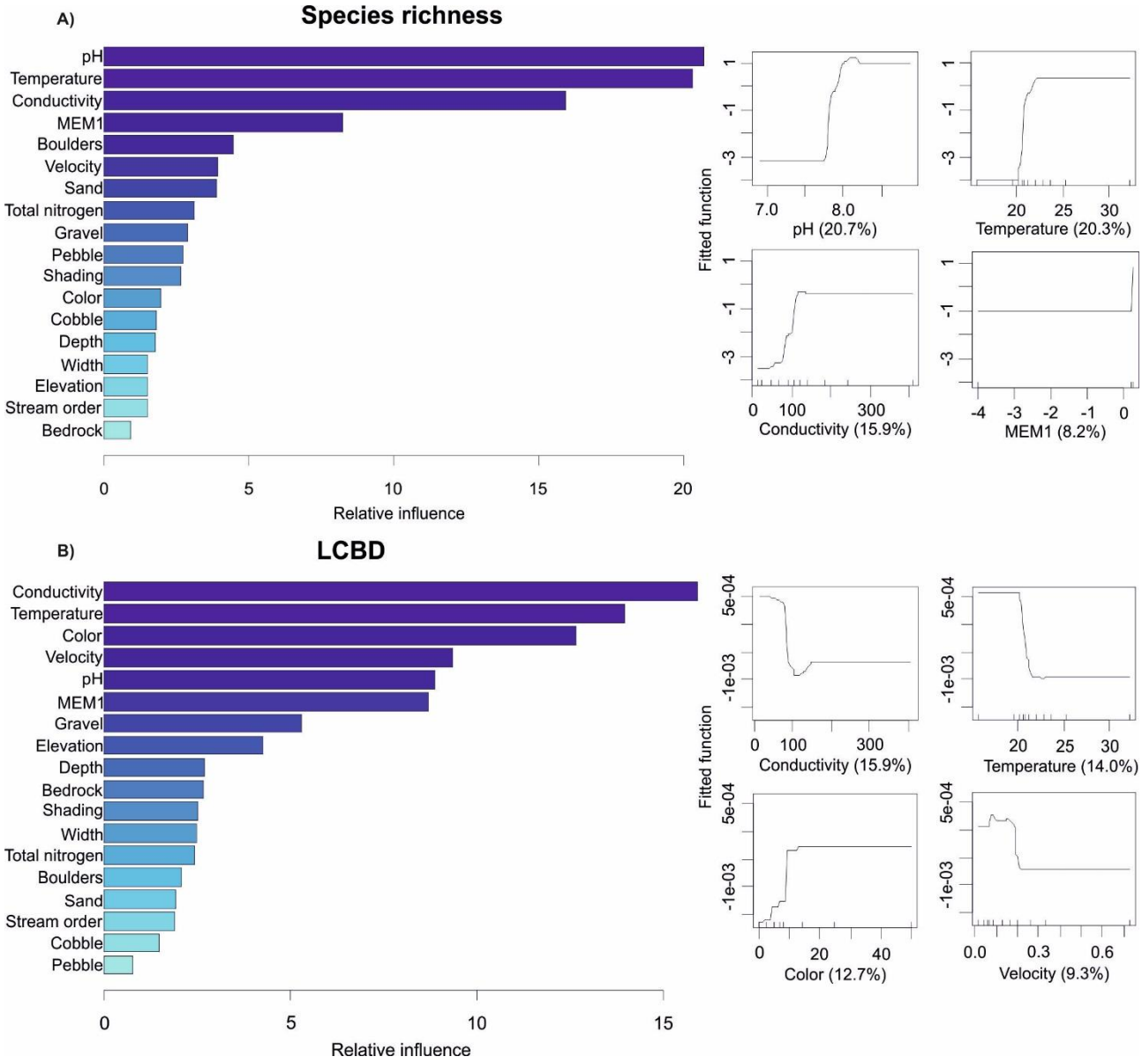
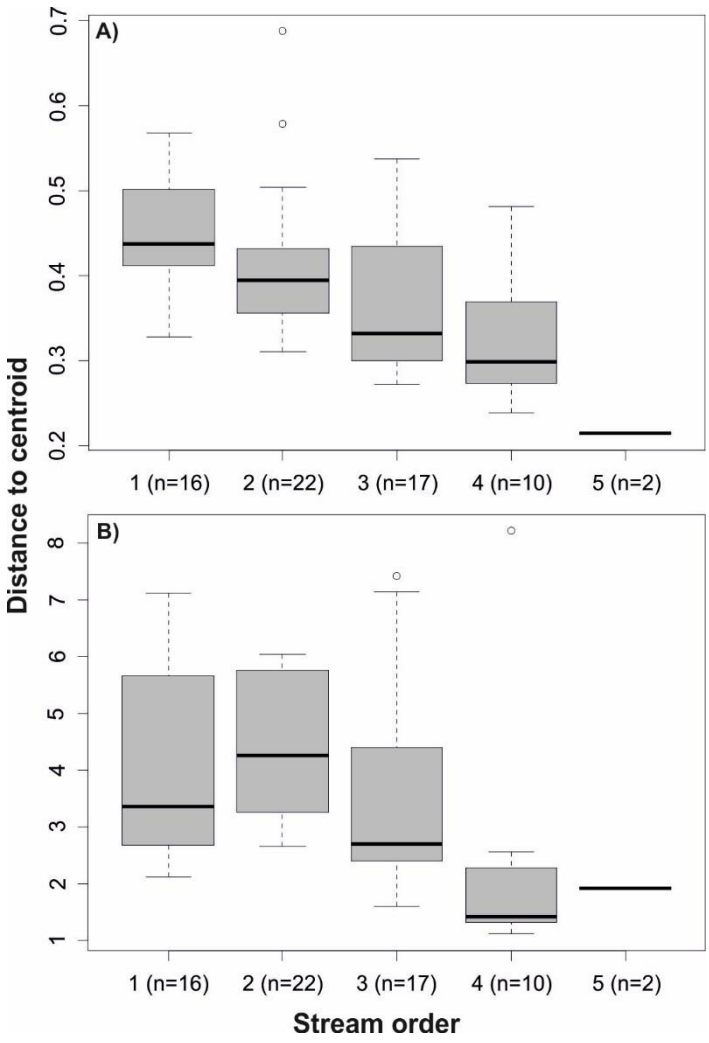


Figure 4 Boosted regression tree summary showing the relative importance of each variable and related fitted function curves for the four most important variables on A) species richness and B) LCBD values. Abbreviations: LCBD (local contribution to beta diversity), MEM1 (spatial eigenvector derived from distance-based Moran´s eigenvector maps)

PERMDISP revealed that there were significant differences in the degree of beta diversity ($P < 0.01$, $F = 6.55$) and environmental heterogeneity ($P < 0.05$, $F = 3.30$) among the stream orders (Q_5, Fig.5a, b). Beta diversity and environmental heterogeneity were higher in headstreams than in higher order streams (Q_5, Fig.5a, b). Albeit based on only five observations, we found a significant positive

398 relationship ($P < 0.05$, $R^2 = 0.736$) between the degree of beta diversity and environmental
399 heterogeneity.



400

401 Figure 5 Boxplots representing the relationship of mean distances from group centroids and stream order based on
402 homogeneity of dispersion analysis. Shown are diatom community compositional data using presence/absence and
403 Sørensen dissimilarity (A) and environmental data using Euclidean distance (B).

404

405 **4. Discussion**

406 The tropical studies focusing on diatom community structure, species richness and beta diversity
407 patterns and their underlying processes are still rare compared with the temperate and boreal studies.
408 In addition, diatom diversity patterns along stream networks are also generally relatively poorly
409 understood. Here, our results shed light on the patterns and controlling factors on these important and
410 understudied issues and may be relevant for watershed management and biomonitoring.

411 Even if largely similar sampling methods and efforts were used, we revealed no clear differences in
412 diatom species richness between tropical and boreal areas. As the sampling area of the boreal dataset
413 was larger than the tropical, and as species richness usually increases with the size of the sampling
414 area, we further tested the difference in species richness between boreal and tropical areas by choosing
415 only the 67 southernmost sites of the boreal dataset in order to have more comparable sampling areas.
416 This approach resulted in total species richness of 298 in boreal data set and the proportion of total
417 species richness on species accumulation curve in 77.9%, which indicates that tropical and boreal
418 streams seem not to differ in diatom species richness. Earlier literature about latitudinal diatom
419 richness patterns have documented very variable richness gradients. Passy (2010) found the
420 relationship of benthic diatom richness and latitude to be U-shaped in a study that comprised running
421 waters of continental U.S.. Highest species richness was found from subtropics and temperate
422 sampling sites and was related with nutrient resources rather than climate. A positive relationship
423 between marine diatom species richness and decreasing latitude was found by Salinas *et al*. (2015),
424 whereas Hillebrand and Azovsky (2001) found no latitudinal effect for diatom species richness and
425 suggested that latitudinal diversity gradient is determined by body size. Hence, our results add to the
426 controversy of the subject and call for further investigation in different types of ecosystems.

427 Our results are congruent with previous studies suggesting that diatom communities are shaped by
428 both environmental and spatial factors (Gothe *et al*., 2013; Heino *et al*., 2012; Heino *et al*., 2014;
429 Potapova and Charles, 2002; Soininen *et al*., 2016; Verleyen *et al*., 2009) indicating both
430 environmental control and dispersal limitation. However, it cannot be excluded that the observed
431 spatial structure here was caused by mass effects (i.e. due to high dispersal rates, a species can be
432 present also in unfavorable sites, (Shmida and Wilson, 1985)) rather than dispersal limitation, due to
433 the relatively small spatial scale (i.e. < 100 km) of the study and high connectivity among sites
434 (Astorga *et al*., 2012). In such conditions, mass effect may be caused by downstream drift of diatom
435 cells (Gothe *et al*., 2013).

436 Among local factors, our study highlighted the importance of water pH and conductivity for diatoms.
437 Conductivity (Astorga *et al*., 2012; Bere and Tundisi, 2011; Biggs, 1995) and water pH (Bere and
438 Tundisi, 2011; Soininen *et al*., 2004; Telford *et al*., 2006) have been repeatedly found to be important
439 controlling factors for diatoms also earlier in wide range of study regions, suggesting that these water
440 chemistry factors are important for diatoms both in tropical and boreal regions. In addition, shading
441 proved to be another important variable structuring diatom communities, a finding, that is further
442 supported by Carpenter and Waite (2000), for example. Thus, although water chemistry plays an
443 indisputable part in structuring diatom communities, it seems that also physical variables may be
444 important, especially light conditions and the substrate of the stream bed. As similar findings emerged
445 in boreal region as well (Jyrkankallio-Mikkola *et al*., 2016; Jyrkänkallio-Mikkola *et al*., 2017), the
446 importance of physical variables in structuring diatom communities applies also to the tropics, which
447 could be important to take into account in biomonitoring.

448 As expected, land use affected diatom communities indirectly through water chemistry and our results
449 suggest that forested high elevation sites including montane cloud forests harbored different diatom
450 communities compared to those found at sites with high conductivity. Similar findings have also been
451 reported by Carpenter and Waite (2000). Thus, including land use variables when unravelling factors
452 controlling the diatom communities may be useful, since this approach provides long-term
453 information of the surrounding habitat in addition to snapshot proximate controlling factors, i.e. water
454 chemistry (Bere and Tundisi, 2011; Jyrkänkallio-Mikkola *et al*., 2017). This finding also emphasizes
455 the importance of montane cloud forests for regional aquatic diversity, and furthermore their
456 conservational importance in general as they are of great importance to regional tree diversity, for
457 example in the Taita Hills (Aerts *et al*., 2011; Schafer *et al*., 2016) and are currently the most
458 threatened and one of the least studied tropical ecosystems (Williams-Linera, 2002).

459 The low explanatory power of direct ordination analyses such as RDA used here is common in
460 microbial studies (Beisner *et al*., 2006; Heino *et al*., 2012; Nabout *et al*., 2009). This might be related

461 to some unmeasured environmental variables although we find this explanation unlikely, since
462 probably the most influential factors structuring diatom communities were included in the analyses.
463 However, we admit that the inclusion of phosphorus might have increased the explanatory power of
464 the model. The surprisingly low phosphorus levels observed might be related to the fact that
465 phosphorus is a major limiting nutrient in old tropical soils (Reich and Oleksyn, 2004) and hence the
466 subsequent leach to the streams may be very low. The low explanatory power might also be due to
467 snap-shot measurements of the local environmental variables, which may not have reflected the
468 prevailing environmental conditions well enough. Indeed, the preceding amount of precipitation may
469 have had an effect on flow conditions, and subsequently, to the diatom community composition.
470 Community assembly mechanisms might also vary for example between generalists and specialists,
471 which may hinder the interpretation of environmental and community relationships (Lindstrom and
472 Langenheder, 2012).

473 As hypothesized, diatom species richness increased with stream order, which is also supported by
474 previous diatom studies (Sherwood *et al*., 2000; Stenger-Kovacs *et al*., 2014). Due to the lack of
475 stream sites above order 5, we cannot rule out the possible decrease in species richness in very large
476 streams and thus, the predicted unimodal pattern of species richness along the river network (RCC
477 theory). However, even if headwater streams harbored lower species richness they comprised more
478 unique communities than higher order stream sites. Headwater streams can be regarded as isolated
479 and less connected habitats compared to downstream sites and hence, the connectivity of the sampling
480 sites may have a strong effect on the community composition via the exchange of individuals (Lopes
481 *et al*., 2014). High uniqueness of headwater assemblages have also been reported for
482 macroinvertebrates (Finn *et al*., 2011), fish (Paller, 1994) and biofilm assemblages (Besemer *et al*.,
483 2013), stressing the conservational importance of headwater streams, contributing strongly to
484 regional gamma and beta diversity. Our results provide evidence that the high uniqueness of
485 headwater communities prevails also in the tropics, perhaps indicating that there are certain general

rules for predicting biodiversity (Finn *et al*., 2011). This result is especially important in areas, where major changes in land use occur together with climate change and consequent draughts may deteriorate or even destroy fragile and important ecosystems such as montane forests. We note that our results are in contrast with the RCC view of headwaters harboring communities less important for stream biodiversity and hence, we think that our results are also important for stream ecology in general.

Agreeing with our findings, similar negative pattern between community uniqueness and species richness across sites has also been found in several biomes and for different organisms; tropical Brazil for dung beetles (da Silva and Hernandez, 2014), subtropical China for plants (Qiao *et al*., 2015), temperate France for phytoplankton (Maloufi *et al*., 2016) and boreal Sweden for macroinvertebrates (Heino *et al*., 2017). Although Legendre and De Caceres (2013) point out that this negative relationship is not general or obligatory (see paper supporting this view for bacteria in the Canadian subarctic by (Comte *et al*., 2016)), it seems to be a relatively common phenomenon at least. Furthermore, this finding suggests that species richness may not be the only or even the best diversity measure for conservational efforts (Dormann *et al*., 2007), whereas community uniqueness might be in some circumstances more effective in this regard. We thus think that the combination of different metrics would probably be most informative approach in many ecosystems.

Of beta diversity components, species turnover was dominating, whereas nestedness was notably smaller. Similar results have been reported also by Wetzel *et al*. (2012) and Piano *et al*. (2017). In our study, it is probable that the high number of singletons contributed to the high turnover rate among sites while richness differences had smaller influence. Total beta diversity and species turnover were higher in headwater streams, whereas nestedness component increased downstream, as we expected. This finding may suggest that the increase of anthropogenic influence downstream changes also the relative effects of mechanisms contributing to beta diversity, thus supporting the results of Gutierrez-Canovas *et al*. (2013). Whereas conductivity, temperature and pH were included into the top five

most significant variables affecting species richness and LCBD, stream order played a minor role. As BRT takes into account interaction effects between variables, this finding suggests that conductivity, pH and water temperature were among the proximate factors causing the differences in diatom community composition between stream orders, as all these variables showed significant positive relationship with the stream order (Appendix D in the supplementary material). Thus, this finding is congruent with Stenger-Kovacs *et al*. (2014) who suggested that the stream order may act as a useful parameter summarizing the physical habitat in streams.

As hypothesized, diatom beta diversity and environmental heterogeneity were higher in headwater streams than in larger stream sites. Thus, our results provided evidence that along with a stronger dispersal limitation in headwaters, environmental heterogeneity may have promoted variation in diatom community composition in headwaters, suggesting that decreasing environmental heterogeneity downstream may at least partly cause subsequent decrease in diatom beta diversity towards larger streams. The higher environmental heterogeneity in headwater streams stresses their conservational importance for regional diversity, which gets support from other studies that have found higher habitat heterogeneity to increase beta diversity for planktonic diatoms (Zorzal-Almeida *et al*., 2017) and phytoplankton in general (Maloufi *et al*., 2016). Although we admit that the conservation of microbial organisms may not have as strong research need as conservation of larger taxa, similar results have also been reported for stream macroinvertebrates (Astorga *et al*., 2014) emphasizing the effect of environmental heterogeneity to regional diversity. Consequently strong emphasis should be placed on headwater stream conservation and management in general.

5. Conclusions

Our results indicate that tropical streams do not seem to harbor more diatom species compared to boreal streams. Studies aiming to unravel the determinants of diatom communities are encouraged to

consider spatial, land use and physical variables in addition to water chemistry variables. Land use variables may affect diatom communities indirectly through water chemistry and can provide long term information of the surrounding habitat. Diatom species richness and the uniqueness of the communities may vary across stream orders with headwater sites encompassing more unique communities and higher order streams higher species richness. This results is important to take into consideration in biomonitoring and studies focusing on stream diatom species richness and community composition. Environmental heterogeneity seems to be higher in headwater streams and also relate with the degree of diatom beta diversity and, thus, our results emphasize the management and conservational importance of headwater streams.

Acknowledgements

We thank the Academy of Finland. The study is done under the following projects funded by Academy of Finland: Scaling biodiversity in tropical and boreal streams: implications for diversity mapping and environmental assessment (grant 273560) and TAITAWATER - Integrated land cover-climate-ecosystem process study for water management in East African highlands, and Finnish Culture Foundation through project Remote sensing of water harvesting and carbon sequestration by forests in the Taita Hills, Kenya. Taita Research Station of the University of Helsinki is acknowledged for logistical support. Research permit NCST/RCD/17/012/33 for Taita Research Station from National Council for Science and Technology of Kenya is greatly acknowledged.

References

Adhikari, H., Heiskanen, J., Siljander, M., Maeda, E., Heikinheimo, V., Pellikka, P., 2017. Determinants of aboveground biomass across an Afromontane landscape mosaic in Kenya. Submitted manuscript.

560 Aerts, R., Thijs, K.W., Lehouck, V., Beentje, H., Bytebier, B., Matthysen, E., Gulinck, H., Lens, L.,
561 Muys, B., 2011. Woody plant communities of isolated Afromontane cloud forests in Taita Hills,
562 Kenya. Plant Ecol 212, 639-649.

563 Al-Shami, S.A., Heino, J., Salmah, M.R.C., Abu Hassan, A., Suhaila, A.H., Madrus, M.R., 2013.
564 Drivers of beta diversity of macroinvertebrate communities in tropical forest streams. Freshwater
565 Biol 58, 1126-1137.

566 Alcantara-Ayala, I., Esteban-Chavez, O., Parrot, J.F., 2006. Landsliding related to land-cover
567 change: A diachronic analysis of hillslope instability distribution in the Sierra Norte, Puebla,
568 Mexico. Catena 65, 152-165.

569 Allan, D.J., Castillo, M.M., 2007. Stream Ecology: Structure and function of running waters, 2nd
570 ed. Springer, P.O.Box 17 3300 AA Dordrecht The Netherlands.

571 Anderson, M.J., 2006. Distance-based tests for homogeneity of multivariate dispersions. Biometrics
572 62, 245-253.

573 Anderson, M.J., Ellingsen, K.E., McArdle, B.H., 2006. Multivariate dispersion as a measure of beta
574 diversity. Ecol Lett 9, 683-693.

575 Astorga, A., Death, R., Death, F., Paavola, R., Chakraborty, M., Muotka, T., 2014. Habitat
576 heterogeneity drives the geographical distribution of beta diversity: the case of New Zealand stream
577 invertebrates. Ecol Evol 4, 2693-2702.

578 Astorga, A., Oksanen, J., Luoto, M., Soininen, J., Virtanen, R., Muotka, T., 2012. Distance decay of
579 similarity in freshwater communities: do macro- and microorganisms follow the same rules? Global
580 Ecol Biogeogr 21, 365-375.

581 Baselga, A., Orme, D., Villeger, S., De Bortoli, J., Leprieur, F., 2017. Partitioning Beta Diversity
582 into Turnover and Nestedness Components. R package ´betapart´ version 1.4-1.

583 Beentje, H.J., Ndiang'ui, N., 1988. An ecological and floristic study of the forests of the Taita Hills,
584 Kenya. Utafiti 1, 23-66.

585 Beisner, B.E., Peres-Neto, P.R., Lindstrom, E.S., Barnett, A., Longhi, M.L., 2006. The role of
586 environmental and spatial processes in structuring lake communities from bacteria to fish. Ecology
587 87, 2985-2991.

588 Bellinger, B.J., Cocquyt, C., O'Reilly, C.M., 2006. Benthic diatoms as indicators of eutrophication
589 in tropical streams. Hydrobiologia 573, 75-87.

590 Benz, U.C., Hofmann, P., Willhauck, G., Lingenfelder, I., Heynen, M., 2004. Multi-resolution,
591 object-oriented fuzzy analysis of remote sensing data for GIS-ready information. Isprs J
592 Photogramm 58, 239-258.

593 Bere, T., 2014. Ecological preferences of benthic diatoms in a tropical river system in Sao Carlos-
594 SP, Brazil. Trop Ecol 55, 47-61.

595 Bere, T., Tundisi, J.G., 2011. Influence of land-use patterns on benthic diatom communities and
596 water quality in the tropical Monjolinho hydrological basin, Sao Carlos-SP, Brazil. Water Sa 37,
597 93-102.

598 Besemer, K., Singer, G., Quince, C., Bertuzzo, E., Sloan, W., Battin, T.J., 2013. Headwaters are
599 critical reservoirs of microbial diversity for fluvial networks. P Roy Soc B-Biol Sci 280.

600 Biggs, B.J.F., 1995. The Contribution of Flood Disturbance, Catchment Geology and Land-Use to
601 the Habitat Template of Periphyton in Stream Ecosystems. Freshwater Biol 33, 419-438.

602 Bojorge-Garcia, M., Carmona, J., Ramirez, R., 2014. Species richness and diversity of benthic
603 diatom communities in tropical mountain streams of Mexico. Inland Waters 4, 279-292.

604 Breiman, L., 2001. Random forests. Mach Learn 45, 5-32.

605 Brink, A.B., Bodart, C., Brodsky, L., Defourney, P., Ernst, C., Donney, F., Lupi, A., Tuckova, K.,
606 2014. Anthropogenic pressure in East Africa-Monitoring 20 years of land cover changes by means
607 of medium resolution satellite data. Int J Appl Earth Obs 28, 60-69.

608 Carpenter, K.D., Waite, I.R., 2000. Relations of habitat-specific algal assemblages to land use and
609 water chemistry in the Willamette Basin, Oregon. Environ Monit Assess 64, 247-257.

610 Clark, B.J.F., Pellikka, P.K.E., 2009. Landscape analysis using multi-scale segmentation and
611 objectoriented classification, in: Röder, A., Hill, J. (Eds.), Recent Advances in Remote Sensing and
612 Geoinformation Processing for Land Degradation Assessment. CRS Press, Taylor & Francis Group,
613 Boca Raton, FL 33487-2742, USA, pp. 323-341.

614 Colwell, R.K., Coddington, J.A., 1994. Estimating Terrestrial Biodiversity through Extrapolation.
615 Philos T Roy Soc B 345, 101-118.

616 Comte, J., Lovejoy, C., Crevecoeur, S., Vincent, W.F., 2016. Co-occurrence patterns in aquatic
617 bacterial communities across changing permafrost landscapes. Biogeosciences 13, 175-190.

618 Concoran, E., Nellermann, C., Baker, E., Bos. R., Osborn, D., Savelli, H.E., 2010. Sick water? The
619 central role of wastewater management in sustainable development. A rapid response assessment.
620 United Nations Environment Programme, UN-HABITAT, GRID-Arendal, Norway.

621 Crump, B.C., Adams, H.E., Hobbie, J.E., Kling, G.W., 2007. Biogeography of bacterioplankton in
622 lakes and streams of an arctic tundra catchment. Ecology 88, 1365-1378.

623 da Silva, P.G., Hernandez, M.I.M., 2014. Local and Regional Effects on Community Structure of
624 Dung Beetles in a Mainland-Island Scenario. Plos One 9.

625 Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carre, G., Marquez, J.R.G., Gruber,
626 B., Lafourcade, B., Leitao, P.J., Munkemuller, T., McClean, C., Osborne, P.E., Reineking, B.,
627 Schroder, B., Skidmore, A.K., Zurell, D., Lautenbach, S., 2013. Collinearity: a review of methods
628 to deal with it and a simulation study evaluating their performance. Ecography 36, 27-46.

629 Dormann, C.F., Schweiger, O., Augenstein, I., Bailey, D., Billeter, R., de Blust, G., DeFilippi, R.,
630 Frenzel, M., Hendrickx, F., Herzog, F., Klotz, S., Liira, J., Maelfait, J.P., Schmidt, T., Speelmans,
631 M., van Wingerden, W.K.R.E., Zobel, M., 2007. Effects of landscape structure and land-use
632 intensity on similarity of plant and animal communities. Global Ecol Biogeogr 16, 774-787.

633 Dray, S., Guillaume, B., Borcard, D., Guenard, G., Jombart, T., Larocque, G., Legendre, P., Madi,
634 M., Wagner, H.H., 2017. Package 'adespatial'. Multivariate multiscale spatial analysis.

635 Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. J Anim
636 Ecol 77, 802-813.

637 Erdogan, H.E., Pellikka, P.K.E., Clark, B., 2011. Modelling the impact of land-cover change on
638 potential soil loss in the Taita Hills, Kenya, between 1987 and 2003 using remote-sensing and
639 geospatial data. Int J Remote Sens 32, 5919-5945.

640 ESA, European Spatial Agency, 2015. Sentinel-2 user handbook. ESA Standard Document. 64.

641 Finn, D.S., Bonada, N., Murria, C., Hughes, J.M., 2011. Small but mighty: headwaters are vital to
642 stream network biodiversity at two levels of organization. J N Am Benthol Soc 30, 963-980.

643 Gaston, K.J., 2000. Global patterns in biodiversity. Nature 405, 220-227.

644 Gereau, R.E., Taylor, C.M., Luke, W.R.Q., 2006. Endemic plant species of the Eastern Arc
645 Mountains of Kenya and Tanzania: analysis and refinement of distributional patterns. In: Taxonomy
646 and Ecology of African Plants, their Conservation and Sustainable Use., in: Ghazanfar, S.A.,
647 Beentje, H.J. (Eds.), Proceedings of the 17th AETFAT Congress, Addis Ababa, Ethiopia. Royal
648 Botanic Gardens, Kew, pp. 267-277.

649 Giraudoux, P., 2015. Data Analysis in Ecology. R package version 1.6.3.

650 Google Earth 7.1.8.3036…

651 Gothe, E., Angeler, D.G., Gottschalk, S., Lofgren, S., Sandin, L., 2013. The Influence of
652 Environmental, Biotic and Spatial Factors on Diatom Metacommunity Structure in Swedish
653 Headwater Streams. Plos One 8.

654 Grant, E.H.C., Lowe, W.H., Fagan, W.F., 2007. Living in the branches: population dynamics and
655 ecological processes in dendritic networks. Ecol Lett 10, 165-175.

656 Green, J.L., Holmes, A.J., Westoby, M., Oliver, I., Briscoe, D., Dangerfield, M., Gillings, M.,
657 Beattie, A.J., 2004. Spatial scaling of microbial eukaryote diversity. Nature 432, 747-750.

658 Gutierrez-Canovas, C., Millan, A., Velasco, J., Vaughan, I.P., Ormerod, S.J., 2013. Contrasting
659 effects of natural and anthropogenic stressors on beta diversity in river organisms. Global Ecol
660 Biogeogr 22, 796-805.

661 Heino, J., Bini, L.M., Andersson, J., Bergsten, J., Bjelke, U., Johansson, F., 2017. Unravelling the
662 correlates of species richness and ecological uniqueness in a metacommunity of urban pond insects.
663 Ecol Indic 73, 422-431.

664 Heino, J., Bini, L.M., Karjalainen, S.M., Mykra, H., Soininen, J., Vieira, L.C.G., Diniz, J.A.F.,
665 2010. Geographical patterns of micro-organismal community structure: are diatoms ubiquitously
666 distributed across boreal streams? Oikos 119, 129-137.

667 Heino, J., Gronroos, M., Soininen, J., Virtanen, R., Muotka, T., 2012. Context dependency and
668 metacommunity structuring in boreal headwater streams. Oikos 121, 537-544.

669 Heino, J., Tolkkinen, M., Pirttila, A.M., Aisala, H., Mykra, H., 2014. Microbial diversity and
670 community-environment relationships in boreal streams. J Biogeogr 41, 2234-2244.

671 Hijmans, R.J.P., S.; Leathwick, J.; Elith, J., 2015. Dismo: Species Distrubution Modelling. R
672 package version 1.0-12.

673 Hill, B.H., Herlihy, A.T., Kaufmann, P.R., Stevenson, R.J., McCormick, F.H., Johnson, C.B., 2000.
674 Use of periphyton assemblage data as an index of biotic integrity. J N Am Benthol Soc 19, 50-67.

675 Hillebrand, H., 2004. On the generality of the latitudinal diversity gradient. Am Nat 163, 192-211.

676 Hillebrand, H., Azovsky, A.I., 2001. Body size determines the strength of the latitudinal diversity
677 gradient. Ecography 24, 251-256.

678 Hohenthal, J., Owidi, E., Minoia, P., Pellikka, P., 2015. Local assessment of changes in water-
679 related ecosystem services and their management: DPASER conceptual model and its application in
680 Taita Hills, Kenya. International Journal of Biodiversity Science, Ecosystem Services &
681 Management 11, 225-238.

682 IUCN, 2009. Wildlife in a changing world and analysis of the 2008 IUCN Red List of Threatened
683 Species, Vié, J.-C., Hilton-Taylor, C., Stuart, S.N. (Eds.), Gland, Switzerland.

684 Januchowski-Hartley, S.R., Pearson, R.G., Puschendorf, R., Rayner, T., 2011. Fresh Waters and
685 Fish Diversity: Distribution, Protection and Disturbance in Tropical Australia. Plos One 6.

686 Jyrkankallio-Mikkola, J., Heino, J., Soininen, J., 2016. Beta diversity of stream diatoms at two
687 hierarchical spatial scales: implications for biomonitoring. Freshwater Biol 61, 239-250.

688 Jyrkankallio-Mikkola, J., Meier, S., Heino, J., Laamanen, T., Pajunen, V., Tolonen, K.T.,
689 Tolkkinen, M., Soininen, J., 2017 Disentangling multi-scale environmental effects on stream
690 microbial communities. J Biogeogr 44, 1512–1523, doi:10.1111/jbi.13002

691 Krammer, K., Lange-Bertalot, H., 1986-1991. Bacillariophyceae. Süβwasserflora von Mitteleuropa
692 1:4, 2. Fischer, Stuttgart.

693 Lawton, J.H., 1999. Are there general laws in ecology? Oikos 84, 177-192.

694 Legendre, P., Borcard, D., Peres-Neto, P.R., 2005. Analyzing beta diversity: Partitioning the spatial
695 variation of community composition data. Ecol Monogr 75, 435-450.

696 Legendre, P., De Caceres, M., 2013. Beta diversity as the variance of community data: dissimilarity
697 coefficients and partitioning. Ecol Lett 16, 951-963.

698 Legendre, P., Gallagher, E.D., 2001. Ecologically meaningful transformations for ordination of
699 species data. Oecologia 129, 271-280.

700 Legendre, P., Legendre, L., 2012. Numerical ecology, 3 ed. Elsevier, Amsterdam.

701 Lindstrom, E.S., Langenheder, S., 2012. Local and regional factors influencing bacterial community
702 assembly. Env Microbiol Rep 4, 1-9.

703 Lopes, P.M., Bini, L.M., Declerck, S.A.J., Farjalla, V.F., Vieira, L.C.G., Bonecker, C.C., Lansac-
704 Toha, F.A., Esteves, F.A., Bozelli, R.L., 2014. Correlates of Zooplankton Beta Diversity in Tropical
705 Lake Systems. Plos One 9.

706 Lovett, J.C., Wasser, S.K., 1993. Biogeography and Ecology of the Rainforests of Eastern Africa.
707 Cambridge University Press, Cambridge.

708 Maloufi, S., Catherine, A., Mouillot, D., Louvard, C., Coute, A., Bernard, C., Troussellier, M.,
709 2016. Environmental heterogeneity among lakes promotes hyper -diversity across phytoplankton
710 communities. Freshwater Biol 61, 633-645.

711 Mangadze, T., Bere, T., Mwedzi, T., 2015. Epilithic diatom flora in contrasting land-use settings in
712 tropical streams, Manyame Catchment, Zimbabwe. Hydrobiologia 753, 163-173.

713 Medley, K.E., J.K. Maingi, 2014. Biogeographic patterns of forest diversity at Mt. Kasigau, Kenya.
714 J East Af Nat Hist Soc Natl Mus 103: 1-24.

715 Meyer, J.L., Strayer, D.L., Wallace, J.B., Eggert, S.L., Helfman, G.S., Leonard, N.E., 2007. The
716 contribution of headwater streams to biodiversity in river networks. J Am Water Resour As 43, 86-
717 103.

718 Nabout, J.C., Siqueira, T., Bini, L.M., Nogueira, I.D., 2009. No evidence for environmental and
719 spatial processes in structuring phytoplankton communities. Acta Oecol 35, 720-726.

720 Oksanen, J., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., O'Hara, R.B., Simpson, G.L.,
721 Solymos, P., Stevens, M.H.H., Wagner, H., 2015. Vegan: Community Ecology Package. R package
722 version 2.3.2.

723 Omoro, L.M.A., Starr, M., Pellikka, P.K.E., 2013. Tree biomass and soil carbon stocks in
724 indigenous forests in comparison to plantations of exotic species in the Taita Hills of Kenya. Silva
725 Fenn 47.

726 Pajunen, V., Luoto, M., Soininen, J., 2016. Climate is an important driver for stream diatom
727 distributions. Global Ecol Biogeogr 25, 198-206.

728 Pajunen, V., Luoto, M., Soininen, J. 2017. Unravelling direct and indirect effects of hierarchical
729 factors driving microbial stream communities. J Biogeogr, In press.

730 Paller, M.H., 1994. Relationships between Fish Assemblage Structure and Stream Order in South-
731 Carolina Coastal-Plain Streams. T Am Fish Soc 123, 150-161.

732 Palmer, M.W., 1990. The Estimation of Species Richness by Extrapolation. Ecology 71, 1195-
733 1198.

734 Passy, S.I., 2010. A distinct latitudinal gradient of diatom diversity is linked to resource supply.
735 Ecology 91, 36-41.

736 Pellikka, P.K.E., Clark, B.J.F., Gonsamo-Gosa, A., Himberg, N., Hurskainen, P., Maeda, E.E.,
737 Mwang´ombe, L.M.A., Siljander, M., 2013. Agricultural expansion and its consequences in the
738 Taita Hills, Kenya, in: Paron, P., Olago, D., Thine Omuto, C. (Eds.), Kenya: a Natural Outlook :
739 Geo-Environmental Resources and Hazards . Developments in Earth Surface Processes. North-
740 Holland Pub.Co, Amsterdam, The Netherlands, pp. 165-179.

741 Pellikka, P.K.E., Lotjonen, M., Sijander, M., Lens, L., 2009. Airborne remote sensing of
742 spatiotemporal change (1955-2004) in indigenous and exotic forest cover in the Taita Hills, Kenya.
743 Int J Appl Earth Obs 11, 221-232.

744 Pellikka, P.K.E., Heikinheimo, V., Hietanen, J., Schäfer, E., Siljander, M., Heiskanen, J..
745 Assessment of the impact of land cover change on aboveground carbon stocks using airborne lidar
746 data and satellite imagery in Taita Hills, Kenya. Submitted manuscript

747 Piano, E., Falasco, E., Bona, F., 2017 How does water scarcity affect spatial and temporal patterns
748 of diatom community assemblages in Mediterranean streams? Freshwater Biol 62:1276–1287, DOI:
749 10.1111/fwb.12944.

750 Pimm, S.L., Russell, G.J., Gittleman, J.L., Brooks, T.M., 1995. The future of biodiversity. Science
751 269, 347-350.

752 Platts, P.J., Burgess, N.D., Gereau, R.E., Lovett, J.C., Marshall, A.R., McClean, C.J., Pellikka,
753 P.K.E., Swetnam, R.D., Marchant, R., 2011. Delimiting tropical mountain ecoregions for
754 conservation. Environ Conserv 38, 312-324.

755 Potapova, M.G., Charles, D.F., 2002. Benthic diatoms in USA rivers: distributions along spatial and
756 environmental gradients. J Biogeogr 29, 167-187.

757 Qiao, X.J., Li, Q.X., Jiang, Q.H., Lu, J.M., Franklin, S., Tang, Z.Y., Wang, Q.G., Zhang, J.X., Lu,
758 Z.J., Bao, D.C., Guo, Y.L., Liu, H.B., Xu, Y.Z., Jiang, M.X., 2015. Beta diversity determinants in
759 Badagongshan, a subtropical forest in central China. Sci Rep-Uk 5.

760 Reich, P.B., Oleksyn, J., 2004. Global patterns of plant leaf N and P in relation to temperature and
761 latitude. P Natl Acad Sci USA 101, 11001-11006.

762 Ridgeway, G., 2013. Generalized boosted regression models. R package version 2.1.1.

763 Rikkinen, J., Laine, T., Pellikka, P. 2015. Water's Journey (A documentary film). Wildheart
764 Productions Oy. Taita Research Station of the University of Helsinki.

765 Rosenzweig, M.L., 1995. Species Diversity in Space and Time. Cambridge University Press,
766 Cambridge.

767 Salinas, H.F.O., Alder, V.A., Puig, A., Boltovskoy, D., 2015. Latitudinal diversity patterns of
768 diatoms in the Southwestern Atlantic and Antarctic waters. J Plankton Res 37, 659-665.

769 Schafer, E., Heiskanen, J., Heikinheimo, V., Pellikka, P., 2016. Mapping tree species diversity of a
770 tropical montane forest by unsupervised clustering of airborne imaging spectroscopy data. Ecol
771 Indic 64, 49-58.

772 Sherwood, A.R., Rintoul, T.L., Muller, K.M., Sheath, R.G., 2000. Seasonality and distribution of
773 epilithic diatoms, macroalgae and macrophytes in a spring-fed stream system in Ontario, Canada.
774 Hydrobiologia 435, 143-152.

775 Shmida, A., Wilson, M.V., 1985. Biological Determinants of Species-Diversity. J Biogeogr 12, 1-
776 20.

777 Socolar, J.B., Gilroy, J.J., Kunin, W.E., Edwards, D.P., 2016. How Should Beta-Diversity Inform
778 Biodiversity Conservation? Trends Ecol Evol 31, 67-80.

779 Soininen, J., 2007. Environmental and spatial control of freshwater diatoms - A review. Diatom Res
780 22, 473-490.

781 Soininen, J., Bartels, P., Heino, J., Luoto, M., Hillebrand, H., 2015. Toward More Integrated
782 Ecosystem Research in Aquatic and Terrestrial Environments. Bioscience 65, 174-182.

783 Soininen, J., Jamoneau, A., Rosebery, J., Passy, S.I., 2016. Global patterns and drivers of species
784 and trait composition in diatoms. Global Ecol Biogeogr 25, 940-950.

785 Soininen, J., Paavola, R., Muotka, T., 2004. Benthic diatom communities in boreal streams:
786 community structure in relation to environmental and spatial gradients. Ecography 27, 330-342.

787 Stam, Å., Enroth, J., Malombe, I., Pellikka, P., Rikkinen, J., 2017. Experimental transplants reveal
788 strong environmental effects on the growth on non-vascular epiphytes in Afromontane Forests.
789 Biotropica, In press

790 Starhler, A.N., 1957 Quantitative Analysis of Watershed Geomorphology. Transactions, American
791 Geophysical, Union 38:6.

792 Stenger-Kovacs, C., Toth, L., Toth, F., Hajnal, E., Padisak, J., 2014. Stream order-dependent
793 diversity metrics of epilithic diatom assemblages. Hydrobiologia 721, 67-75.

794 Strayer, D.L., Dudgeon, D., 2010. Freshwater biodiversity conservation: recent progress and future
795 challenges. J N Am Benthol Soc 29, 344-358.

796 Taylor, J.C., Harding, W.R., Archibald, C.G.M., 2007. An illustrated guide to some common
797 diatom species from South Africa, Development of a Diatom Assessment Protocol (DAP) for River
798 Health Assessment. Water Research Comission, Republic of South Africa, p. 178.

799 Telford, R.J., Vandvik, V., Birks, H.J.B., 2006. Dispersal limitations matter for microbial
800 morphospecies. Science 312, 1015-1015.

801 Thomas, C.D., Cameron, A., Green, R.E., Bakkenes, M., Beaumont, L.J., Collingham, Y.C.,
802 Erasmus, B.F.N., de Siqueira, M.F., Grainger, A., Hannah, L., Hughes, L., Huntley, B., van
803 Jaarsveld, A.S., Midgley, G.F., Miles, L., Ortega-Huerta, M.A., Peterson, A.T., Phillips, O.L.,
804 Williams, S.E., 2004. Extinction risk from climate change. Nature 427, 145-148.

805 Tonkin, J.D., Arimoro, F.O., Haase, P., 2016. Exploring stream communities in a tropical
806 biodiversity hotspot: biodiversity, regional occupancy, niche characteristics and environmental
807 correlates. Biodivers Conserv 25, 975-993.

808 UNESCO, 2009. World Water Assessment Programme, Water in a Changing World. The United
809 Nations World Water Development Report 3, UNESCO.

810 US- Environmental Protection Agency, 2000. The quality of Our Nation's Waters, EPA-841-R-02-
811 001, US EPA, Wasington DC.

812 Wang, Y.K., Stevenson, R.J., Metzmeier, L., 2005. Development and evaluation of a diatom-based
813 index of Biotic Integrity for the Interior Plateau Ecoregion, USA. J N Am Benthol Soc 24, 990-
814 1008.

815 Vannote, R.L.M., G.W.; Cummins, K.W.; Sedell, J.R.; Cushing, C.E., 1980. The river continuum
816 concept. Can J Fish Aquat Sci 37, 130-137.

817 Wei, T., Simko, V., 2016. Corrplot: Visualization of a Correlation Matrix. R package version 0.77.

818 Wentworth, C.K., 1922. A Scale of Grade and Class Terms for Clastic Sediments. The Journal of
819 Geology 30, 377-392.

820 Verleyen, E., Vyverman, W., Sterken, M., Hodgson, D.A., De Wever, A., Juggins, S., Van de
821 Vijver, B., Jones, V.J., Vanormelingen, P., Roberts, D., Flower, R., Kilroy, C., Souffreau, C.,
822 Sabbe, K., 2009. The importance of dispersal related and local factors in shaping the taxonomic
823 structure of diatom metacommunities. Oikos 118, 1239-1249.

824 Wetzel, C.E., Bicudo, D.D., Ector, L., Lobo, E.A., Soininen, J., Landeiro, V.L., Bini, L.M., 2012.
825 Distance Decay of Similarity in Neotropical Diatom Communities. Plos One 7.

826 Wiens, J.J., 2016. Climate-Related Local Extinctions Are Already Widespread among Plant and
827 Animal Species. Plos Biol 14.

828 Williams-Linera, G., 2002. Tree species richness complementarity, disturbance and fragmentation
829 in a Mexican tropical montane cloud forest. Biodivers Conserv 11, 1825-1843.

830 Vinson, M.R., Hawkins, C.P., 1998. Biodiversity of stream insects: Variation at local, basin, and
831 regional scales. Annu Rev Entomol 43, 271-293.

832 Vorosmarty, C.J., McIntyre, P.B., Gessner, M.O., Dudgeon, D., Prusevich, A., Green, P., Glidden,
833 S., Bunn, S.E., Sullivan, C.A., Liermann, C.R., Davies, P.M., 2010. Global threats to human water
834 security and river biodiversity (vol 467, pg 555, 2010). Nature 468, 334-334.

835 Zeglin, L.H., 2015. Stream microbial diversity in response to environmental changes: review and
836 synthesis of existing research. Front Microbiol 6.

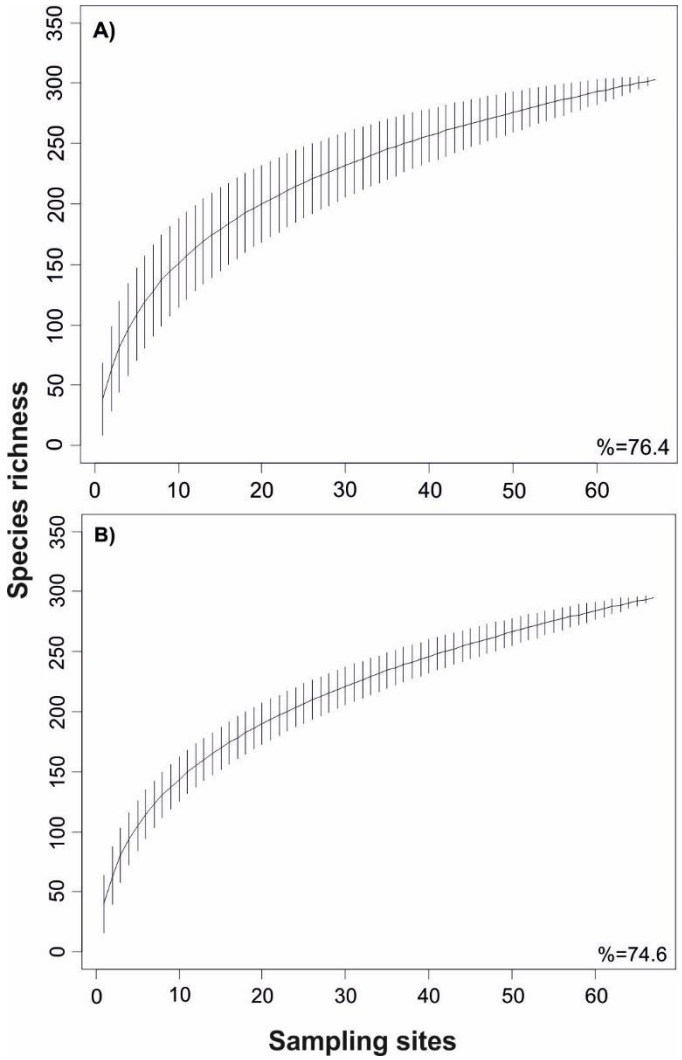
837 Zorzal-Almeida, S., Bini, L.M., Bicudo, D.C., 2017 Beta diversity of diatoms is driven by
838 environmental heterogeneity, spatial extent and productivity. Hydrobiologia, In press,
839 doi:10.1007/s10750-017-3117-3

840

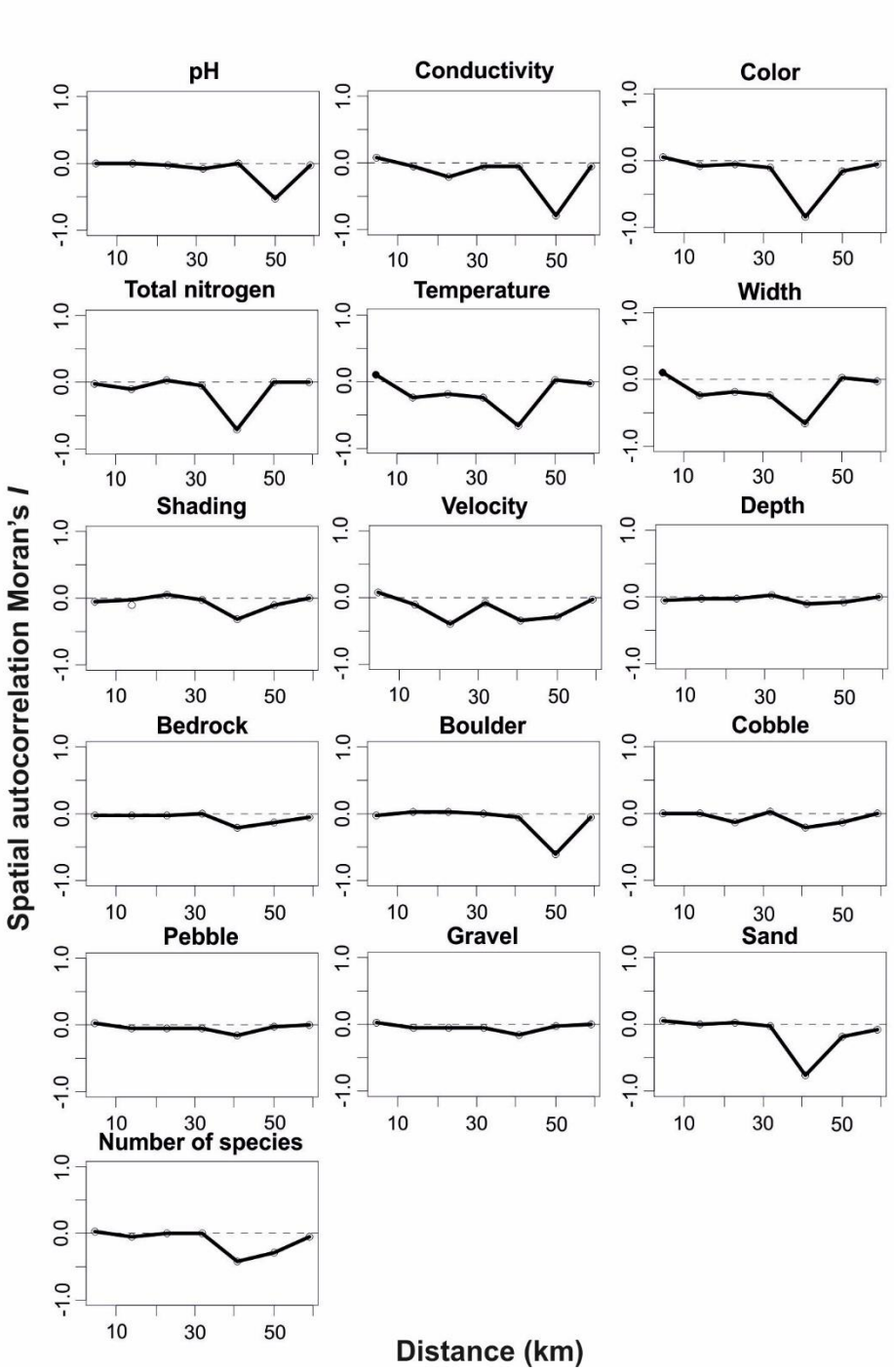
841 Appendix A List of measured environmental and land use variables (average values and ranges). * removed from
842 multivariate analysis due to low concentrations, **removed from further analysis to avoid 100 % coverage of the
843 variable group.

	Variable	Mean values	Range
Environmental variables (n=67)	Elevation (m)	1293.00	(713.00 – 1929.00)
	pH	7.90	(6.90 – 8.89)
	Conductivity (mSm^{-1})	120.28	(15.10- 408.70)
	Colour (Pt mgl^{-1})	9.02	(0 - 50)
	Total P(μgl^{-1}) *	14.46	(<10.00 – 69.00)
	Total N (μgl^{-1})	698.06	(150.00 – 2380.00)
	Shading (%)	40.11	(0.00 – 96.30)
	Depth (m)	0.07	(0.01 - 0.23)
	Width (cm)	148.01	(17.00 – 1310.00)
	Velocity (ms^{-1})	0.17	(0.02 - 0.73)
	Sand (%)	46.17	(0.00 - 100.00)
	Gravel (%)	5.29	(0.00 – 46.50)
	Pebble (%)	4.66	(0.00 – 35.00)
	Cobble (%)	7.34	(0.00 – 49.50)
	Boulders (%)	9.11	(0.00 – 64.00)
	Moss (%)**	5.73	(0.00 – 90.00)
Land use variables (n=39)	Catchment area (km^2)	2.32	(0.14 – 25.58)
	Cropland (%)	38.00	(0.00 – 77.76)
	Shrubland (%)	1.93	(0.00 – 23.81)
	Thicket (%)	5.75	(0.00 – 39.14)
	Woodland (%)	21.10	(0.00 – 51.42)
	Plantation forest (%)	15.75	(0.00 – 41.56)
	Broadleaved forest (%)	14.70	(0.00 – 93.00)
	Bare soil and built areas (%)	1.10	(0.00 – 14.00)
	Bare rock (%)	1.50	(0.00 – 15.60)
	Water bodies (%)**	0.01	(0.00 – 0.41)

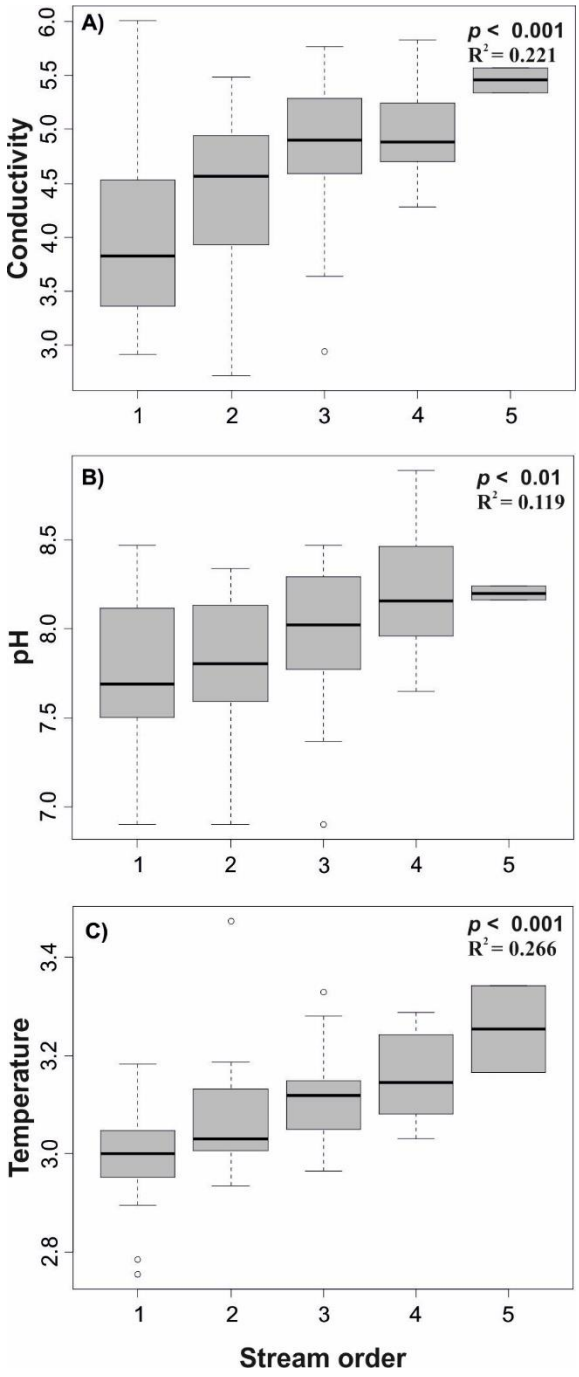
844



Appendix B Species accumulation curves for A) boreal and B) tropical streams. The proportion of estimated species richness that 67 sampling sites covered is shown on the lower right corner.



Appendix C Moran's I correlogram representing spatial autocorrelation for 15 environmental variables and diatom species richness. Black filled circles mark significant values of Moran's I and represent significant positive autocorrelation.



Appendix D Boxplots representing the relationship of A) log transformed conductivity, B) pH, C) log transformed temperature and stream order.